

Published in final edited form as:

Med Image Anal. 2012 December ; 16(8): 1477–1489. doi:10.1016/j.media.2012.06.007.

Class-Specific Weighting for Markov Random Field Estimation: Application to Medical Image Segmentation

James P. Monaco^{*,a} and Anant Madabhushi^{*,a}

^aDepartment of Biomedical Engineering, Rutgers University, 599 Taylor Road, Piscataway, NJ, USA

Abstract

Many estimation tasks require Bayesian classifiers capable of adjusting their performance (e.g. sensitivity/Specificity). In situations where the optimal classification decision can be identified by an exhaustive search over all possible classes, means for adjusting classifier performance, such as probability thresholding or weighting the *a posteriori* probabilities, are well established. Unfortunately, analogous methods compatible with Markov random fields (i.e. large collections of dependent random variables) are noticeably absent from the literature. Consequently, most Markov random field (MRF) based classification systems typically restrict their performance to a single, static operating point (i.e. a paired sensitivity/Specificity). To address this deficiency, we previously introduced an extension of maximum posterior marginals (MPM) estimation that allows certain classes to be weighted more heavily than others, thus providing a means for varying classifier performance. However, this extension is not appropriate for the more popular maximum *a posteriori* (MAP) estimation. Thus, a strategy for varying the performance of MAP estimators is still needed. Such a strategy is essential for several reasons: 1) the MAP cost function may be more appropriate in certain classification tasks than the MPM cost function, 2) the literature provides a surfeit of MAP estimation implementations, several of which are considerably faster than the typical Markov Chain Monte Carlo methods used for MPM, and 3) MAP estimation is used far more often than MPM. Consequently, in this paper we introduce multiplicative weighted MAP (MWMAP) estimation — achieved via the incorporation of multiplicative weights into the MAP cost function — which allows certain classes to be preferred over others. This creates a natural bias for Specific classes, and consequently a means for adjusting classifier performance. Similarly, we show how this multiplicative weighting strategy can be applied to the MPM cost function (in place of the strategy we presented previously), yielding multiplicative weighted MPM (MWMPPM) estimation. Furthermore, we describe how MWMAP and MWMPPM can be implemented using adaptations of current estimation strategies such as iterated conditional modes and MPM Monte Carlo. To illustrate these implementations, we first integrate them into two separate MRF-based classification systems for detecting carcinoma of the prostate (CaP) on 1) digitized histological sections from radical prostatectomies and 2) T2-weighted 4 Tesla *ex vivo* prostate MRI. To highlight the extensibility of MWMAP and MWMPPM to estimation tasks involving more than two classes, we also incorporate these estimation criteria into a MRF-based classifier used to segment synthetic brain MR images. In the context of these tasks, we show how our novel estimation criteria can be used to arbitrarily adjust the sensitivities of these systems, yielding receiver operator characteristic curves (and surfaces).

© 2012 Elsevier B.V. All rights reserved.

^{*}Corresponding authors: jmonaco@rci.rutgers.edu (James P. Monaco), anantm@rci.rutgers.edu (Anant Madabhushi).

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Keywords

Markov Random Fields; Prostate Cancer Detection; Histology; Digital Pathology; Magnetic Resonance Imaging; Maximum Posterior Marginals; Maximum a Posteriori

1. Introduction

The ability to classify multiple objects (e.g. pixels in an image) simultaneously is essential for certain estimation tasks. Within a Bayesian framework, each object (“site” in MRF nomenclature) is modeled as a random variable, and the collection of these random variables (under minor conditions) is called a Markov random field (MRF). If the random variables are assumed independent, we can estimate each in isolation. This estimation typically involves an exhaustive search. For example, obtaining the maximum *a posteriori* (MAP) estimate (of a single random variable) entails calculating the *a posteriori* probability for each possible class, and then choosing the class with the largest probability. However, if the random variables are not independent, the entire MRF must be estimated collectively. Since the number of possible states of the random field is prohibitively large, the exhaustive approach becomes untenable¹. Consequently, more sophisticated schemes, such as relaxation procedures (Geman and Geman, 1984; Besag, 1986), Monte Carlo methods (Marroquin et al., 1987), loopy belief propagation (Yedidia et al., 2000), and graph cuts (Boykov et al., 2001), become necessary. For a comparison of different estimation procedures see (Dubes et al., 1990; Szeliski et al., 2008).

The capability of adjusting classifier performance (e.g. sensitivity/Specificity) with respect to Specific classes is essential for many applications — especially in medical imaging. For example, misclassifying a malignant lesion is typically more egregious than misclassifying a benign lesion. In situations where the optimal classification decision can be identified by an exhaustive search, means for modifying classifier performance, such as probability thresholding or weighting the *a posteriori* probabilities, are well established (Duda et al., 2001). Unfortunately, analogous methods compatible with MRFs are noticeably absent in the literature. Consequently, most MRF-based classification systems restrict their performance to a single, static operating point (i.e. a paired sensitivity/Specificity).

Though MRFs are pervasive in the computer vision and medical imaging literature — addressing such tasks such as segmentation (Pappas, 1992; Farag et al., 2006; Awate et al., 2006; Liu et al., 2009; Scherrer et al., 2009; Marroquin et al., 2002; Bouman and Shapiro, 1994), denoising (Besag, 1986; Figueiredo and Leita, 1997), and texture synthesis (Paget and Longstaff, 1998; Zalesny and Gool, 2001) — relatively little work has discussed means for varying the performance of MRF-based classification systems. To address this deficiency, we recently presented a generalization — and novel Markov Chain Monte Carlo realization — of the maximum posterior marginals (MPM) estimation criterion (Monaco and Madabhushi, 2011). This generalization allows certain classes to be weighted more heavily than others, thus providing a means for varying MRF-based classifier performance. Like all Bayesian estimation techniques, MPM estimation is derived by minimizing the expected value of an underlying cost function. For a given state of the random field, the MPM cost function is simply the sum of the number of sites that are misclassified. To generalize this cost function we weighted each misclassification by the type of classification error. Please see (Monaco and Madabhushi, 2011) for details.

¹If a random field contains N random variables, each of which can assume one of L classes, the total number of possible states is L^N .

However, this approach is Specific to MPM estimation, and is not appropriate for generalizing the very different MAP cost function. For unlike the MPM cost function that sums the site-wise misclassifications, the MAP cost function yields an identical cost for any number of misclassifications greater than zero. Thus, a strategy for varying the performance of MAP estimators is still needed. Such a strategy is essential for several reasons: 1) the MAP cost function may be more appropriate in certain classification tasks than the MPM cost function, 2) the literature provides a surfeit of MAP estimation implementations (Boykov et al., 2001; Dubes et al., 1990; Szeliski et al., 2008; Geman and Geman, 1984; Besag, 1986), several of which (e.g. graph cuts and ICM) are considerably faster than the typical Markov Chain Monte Carlo methods used for MPM (Marroquin et al., 1987; Monaco and Madabhushi, 2011), and 3) MAP estimation is far more prevalent than MPM. (Note that MAP and MPM should not be confused with the algorithms used to implement them, e.g. simulated annealing (Geman and Geman, 1984) for MAP and Markov Chain Monte Carlo (Marroquin et al., 1987) for MPM.)

Aside from our own work, we are aware of very few articles that discuss means for adjusting MRF-based classifier performance, whether using MPM or MAP estimation. In (Comer and Delp, 1999) Comer and Delp employed a Markov prior with a term that applied costs to specified classes. However, they did not discuss this term with respect to varying system performance, nor did they present it within the rigorous framework of Bayesian cost analysis (as we do here). In a seminal paper by Besag (Besag, 1986), the author suggested leveraging a unique property of iterated conditional modes (ICM) to adjust classification results. ICM is an iterative, deterministic procedure that converges to a local maximum of the *a posteriori* probability of a MRF. ICM requires the initial state of the MRF from which to begin the iteration; the choice of this state determines the local maximum to which ICM converges. Thus, varying the initial conditions can vary the classification results. However, the different modes of the *a posteriori* probability (to which ICM converges) do not necessarily correspond to meaningful classifications. Thus, this method, though intuitively appealing, lacks mathematical justification (from the perspective of Bayesian cost).

In this paper, we introduce a generalization of the MAP estimation criteria and present means for its implementation. Specifically, we demonstrate how multiplicative weights can be incorporated into the Bayesian cost function that leads to MAP estimation, thereby biasing the posterior probability to favor certain classes. We refer to estimation using this new cost function as multiplicative weighted MAP (MWMAP) estimation. Coincidentally, the same multiplicative weighting can also be applied to the MPM cost function, yielding multiplicative weighted MPM (MWMPM) estimation. In addition to introducing these novel estimation criteria, we demonstrate how they can be implemented by modifying existing estimation schemes (e.g. ICM). Furthermore, we highlight their significance by incorporating them into two separate classification systems based on MRFs: 1) a system for detecting cancerous glands in histological sections (HSs) from radical prostatectomies (Monaco et al., 2009a) and 2) a system for detecting cancerous regions in T2-weighted 4 Tesla *ex vivo* MRI of the prostate (Viswanath et al., 2012). Over a cohort of 40 digitized HSs and 15 2D MRI sections, respectively, we illustrate how MWMAP and MWMPM estimation can be used to vary classification performance, enabling the construction of receiver operator characteristic (ROC) curves. Finally, to demonstrate the extensibility of MWMAP and MWMPM to estimation tasks involving more than two classes, we incorporate these criteria into a MRF-based classifier used to segment synthetic brain MR images. In this instance, varying classifier performance results in ROC surfaces.

To summarize, the primary contributions of this work are as follows:

- By incorporating class Specific weights into the MAP and MPM estimation criteria, we introduce two novel estimation criteria capable of adjusting the performance of MRF-based classification systems.
- We develop implementations of these criteria by modifying existing realizations of MAP and MPM estimation (e.g. ICM and MPM Monte Carlo).
- We integrate the MWMAP and MWMPM estimation criteria into three MRF-based classification systems, demonstrating the ability of each criterion to arbitrarily adjust system performance.

The remainder of the paper is organized as follows: In Section 2 we review the Bayesian estimation of MRFs, deriving the MAP and MPM estimation criteria using Bayesian risk analysis. Section 3 discusses how class-Specific weights can be incorporated into MAP and MPM, yielding MWMAP and MWMPM. In Section 4 we present implementations of these novel criteria by modifying current MRF-compatible estimation strategies. In Section 5 we present experiments to demonstrate how the estimation criteria can be used to adjust the performance of our systems for detecting CaP (in HSs and MR images) and segmenting MR brain phantoms. Results of these experiments are provided in Section 6. Section 7 offers concluding remarks.

2. Review of Random Fields and Bayesian Risk

2.1. Random Field Definitions and Notation

Let the set $S = \{1, 2, \dots, N\}$ reference N sites to be classified. Each site $s \in S$ has two associated random variables: $X_s \in \Lambda \equiv \{\omega_1, \omega_2, \dots, \omega_L\}$ indicating its state (class) and $Y_s \in \mathbb{R}^D$ representing its D -dimensional feature vector. Particular instances of X_s and Y_s are denoted by the lowercase variables $x_s \in \Lambda$ and $y_s \in \mathbb{R}^D$. Let $\mathbf{X} = (X_1, X_2, \dots, X_N)$ and $\mathbf{Y} = (Y_1, Y_2, \dots, Y_N)$ refer to all random variables X_s and Y_s in aggregate. The state spaces of \mathbf{X} and \mathbf{Y} are the Cartesian products $\Omega = \Lambda^N$ and $\mathbb{R}^{D \times N}$. Instances of \mathbf{X} and \mathbf{Y} are denoted by the lowercase variables $\mathbf{x} = (x_1, x_2, \dots, x_N) \in \Omega$ and $\mathbf{y} = (y_1, y_2, \dots, y_N) \in \mathbb{R}^{D \times N}$. See Table 1 for a list and description of the commonly used notations and symbols in this paper.

Let $G = \{S, E\}$ establish an undirected graph structure on the sites, where S and E are the vertices (sites) and edges, respectively. A neighborhood η_s is the set containing all sites that share an edge with s , i.e. $\eta_s = \{r : r \in S, r \neq s, \{r, s\} \in E\}$. The random field \mathbf{X} is a Markov random field if its local conditional probability functions satisfy the Markov property: $P(X_s = x_s | \mathbf{X}_{-s} = \mathbf{x}_{-s}) = P(X_s = x_s | \mathbf{X}_{\eta_s} = \mathbf{x}_{\eta_s})$, where $\mathbf{x}_{-s} = (x_1, \dots, x_{s-1}, x_{s+1}, \dots, x_N)$, $\mathbf{x}_{\eta_s} = (x_{\eta_s(1)}, \dots, x_{\eta_s(|\eta_s|)})$, and $\eta_s(i) \in S$ is the i^{th} element of the set η_s . Note that in places where it does not create ambiguity, we will simplify the probabilistic notations by omitting the random variables, e.g. $P(\mathbf{x}) \equiv P(\mathbf{X}=\mathbf{x})$.

2.2. Bayesian Risk

Given an observation of the feature vectors \mathbf{Y} , we would like to estimate the states \mathbf{X} . Bayesian estimation advocates selecting the estimate $\hat{\mathbf{x}} \in \Omega$ that minimizes the conditional risk (expected cost)(Duda et al., 2001)

$$R(\mathbf{X}|\hat{\mathbf{x}}, \mathbf{y}) = E[C(\mathbf{X}, \hat{\mathbf{x}})|\mathbf{y}] = \sum_{\mathbf{x} \in \Omega} C(\mathbf{x}, \hat{\mathbf{x}})P(\mathbf{x}|\mathbf{y}), \quad (1)$$

where E indicates expected value and $C(\mathbf{x}, \hat{\mathbf{x}})$ is the cost of selecting labels $\hat{\mathbf{x}}$ when the true labels are \mathbf{x} . In the following subsections we will consider the two most prevalent cost functions used with MRFs.

2.3. Maximum a Posteriori Estimation

The most ubiquitous cost function (though this cost is rarely expressed explicitly) is

$$C_{\text{MAP}}(\mathbf{x}, \widehat{\mathbf{x}}) = 1 - \prod_{s \in S} \delta(x_s - \widehat{x}_s), \quad (2)$$

where δ is the Kronecker delta. Thus, a cost of 1 is incurred if one or more sites are labeled incorrectly. Inserting (2) into (1) yields

$$\begin{aligned} R_{\text{MAP}}(\mathbf{X}|\widehat{\mathbf{x}}, \mathbf{y}) &= \sum_{\mathbf{x} \in \Omega} \left[1 - \prod_{s \in S} \delta(x_s - \widehat{x}_s) \right] P(\mathbf{x}|\mathbf{y}) \\ &= 1 - P(\widehat{\mathbf{x}}|\mathbf{y}). \end{aligned} \quad (3)$$

Minimizing (3) over $\widehat{\mathbf{x}}$ is equivalent to maximizing $P(\widehat{\mathbf{x}}|\mathbf{y})$. Thus, we have maximum *a posteriori* (MAP) estimation, which advocates selecting the $\widehat{\mathbf{x}}$ that maximizes the *a posteriori* probability.

2.4. Maximum Posterior Marginals

As an alternative to MAP estimation, Marroquin *et al.* (Marroquin et al., 1987) suggested using the following cost function

$$C_{\text{MPM}}(\mathbf{x}, \widehat{\mathbf{x}}) = \sum_{s \in S} [1 - \delta(x_s - \widehat{x}_s)]. \quad (4)$$

This function counts the number of sites in $\widehat{\mathbf{x}}$ that are labeled incorrectly. Inserting (4) into (1) yields

$$\begin{aligned} R_{\text{MPM}}(\mathbf{X}|\widehat{\mathbf{x}}, \mathbf{y}) &= \sum_{\mathbf{x} \in \Omega} \left(\sum_{s \in S} [1 - \delta(x_s - \widehat{x}_s)] \right) P(\mathbf{x}|\mathbf{y}) \\ &= \sum_{s \in S} \sum_{\mathbf{x} \in \Omega} P(\mathbf{x}|\mathbf{y}) - \sum_{s \in S} \sum_{\mathbf{x} \in \Omega} \delta(x_s - \widehat{x}_s) P(\mathbf{x}|\mathbf{y}) \\ &= |S| - \sum_{s \in S} P(\widehat{x}_s|\mathbf{y}). \end{aligned} \quad (5)$$

The distributions $P(\widehat{x}_s|\mathbf{y})$ are called the posterior marginals. Minimizing (5) over $\widehat{\mathbf{x}}$ is equivalent to independently maximizing each of these posterior marginals with respect to its corresponding \widehat{x}_s . Hence, this estimation criterion is termed maximum posterior marginals (MPM).

3. Class Weighted Cost Functions

Both the MAP and MPM cost functions weight each class equally. That is, the cost accrued from misclassification is identical across all classes. In this section we introduce generalizations of the MAP and MPM cost functions which provide a means for weighting certain classes more heavily than others. Specifically, we present the multiplicative weighted MAP (MWMAP) and MPM (MWMPM) estimation criteria.

3.1. Multiplicative Weighted Maximum a Posteriori (MWMAP) Estimation

To introduce the ability to favor Specific classes, C_{MAP} can be generalized as follows:

$$C_{\text{MWMAP}}(\mathbf{x}, \widehat{\mathbf{x}}) = \alpha(\mathbf{x}) \left[1 - \prod_{s \in \mathcal{S}} \delta(x_s - \widehat{x}_s) \right], \quad (6)$$

where $\alpha(\mathbf{x}) = \prod_{s \in \mathcal{S}} \alpha(x_s)$ and $\alpha: \Lambda \rightarrow \mathbb{R}_0^+$ are the class dependent weighting functions. Thus, if the estimate $\widehat{\mathbf{x}}$ contains any erroneous labels it accrues a cost of $\alpha(\mathbf{x})$. Note that $\alpha(\mathbf{x})$ can not assume any arbitrary functional form, but is restricted to the product of the independent weighting functions $\alpha(x_s)$. This restriction is necessary for the tractability of subsequent derivations.

Before proceeding we introduce a Definition that will prove useful. Let $\tilde{\mathbf{X}}$ be a random field that differs from \mathbf{X} only with respect to the following probability measure:

$$P(\tilde{\mathbf{X}} = \mathbf{x}) = \frac{1}{Z_\alpha} \alpha(\mathbf{x}) P(\mathbf{X} = \mathbf{x}), \quad (7)$$

where $Z_\alpha = \sum_{\mathbf{x} \in \Omega} \alpha(\mathbf{x}) P(\mathbf{X} = \mathbf{x})$ is the normalizing constant. Note also that

$$P(\tilde{\mathbf{X}} = \mathbf{x} | \mathbf{y}) = \frac{1}{Z_\alpha} \alpha(\mathbf{x}) P(\mathbf{x} | \mathbf{y}). \quad (8)$$

We now insert (6) into (1) yielding

$$\begin{aligned} R_{\text{MWMAP}}(\mathbf{X} | \widehat{\mathbf{x}}, \mathbf{y}) &= \sum_{\mathbf{x} \in \Omega} \alpha(\mathbf{x}) \left[1 - \prod_{s \in \mathcal{S}} \delta(x_s - \widehat{x}_s) \right] P(\mathbf{x} | \mathbf{y}) \\ &= \sum_{\mathbf{x} \in \Omega} \left[1 - \prod_{s \in \mathcal{S}} \delta(x_s - \widehat{x}_s) \right] \alpha(\mathbf{x}) P(\mathbf{x} | \mathbf{y}) \\ &= \sum_{\mathbf{x} \in \Omega} \left[1 - \prod_{s \in \mathcal{S}} \delta(x_s - \widehat{x}_s) \right] Z_\alpha \cdot P(\tilde{\mathbf{X}} = \mathbf{x} | \mathbf{y}) \\ &= Z_\alpha \cdot R_{\text{MAP}}(\tilde{\mathbf{X}} | \widehat{\mathbf{x}}, \mathbf{y}). \end{aligned} \quad (9)$$

Thus, minimizing $R_{\text{MWMAP}}(\mathbf{X} | \widehat{\mathbf{x}}, \mathbf{y})$ is equivalent to minimizing $R_{\text{MAP}}(\tilde{\mathbf{X}} | \widehat{\mathbf{x}}, \mathbf{y})$; and consequently, the optimal labeling is the MAP estimate of $\tilde{\mathbf{X}}$. Since, as shown in (8), the *a posteriori* probability of $\tilde{\mathbf{X}}$ corresponds to the weighted *a posteriori* probability of \mathbf{X} , we refer to this type of estimation as multiplicative weighted MAP estimation. Note that if $\alpha(x_s) \equiv 1$, MWMAP estimation reduces to MAP estimation.

3.2. Multiplicative Weighted Maximum Posterior Marginals (MWMPM)

Class-Specific weights can be similarly incorporated into the MPM cost function:

$$C_{\text{MWMPM}}(\mathbf{x}, \widehat{\mathbf{x}}) = \alpha(\mathbf{x}) \sum_{s \in \mathcal{S}} [1 - \delta(x_s - \widehat{x}_s)]. \quad (10)$$

Mislabeled a site whose true label is x_s has an associated cost of $\alpha(\mathbf{x})$. Consequently, the penalty for mislabeling the single site s depends upon the true labels of all sites $s \in \mathcal{S}$. Inserting (10) into (1) yields

$$\begin{aligned}
R_{\text{MWMPM}}(\mathbf{X}|\widehat{\mathbf{x}}, \mathbf{y}) &= \sum_{\mathbf{x} \in \Omega} \left[\alpha(\mathbf{x}) \prod_{s \in S} [1 - \delta(x_s - \widehat{x}_s)] \right] P(\mathbf{x}|\mathbf{y}) \\
&= \sum_{\mathbf{x} \in \Omega} \left[\prod_{s \in S} [1 - \delta(x_s - \widehat{x}_s)] \right] \alpha(\mathbf{x}) P(\mathbf{x}|\mathbf{y}) \\
&= \sum_{\mathbf{x} \in \Omega} \left[\prod_{s \in S} [1 - \delta(x_s - \widehat{x}_s)] \right] Z_\alpha \cdot P(\tilde{\mathbf{X}}=\mathbf{x}|\mathbf{y}) \\
&= Z_\alpha \cdot R_{\text{MPM}}(\tilde{\mathbf{X}}|\widehat{\mathbf{x}}, \mathbf{y}).
\end{aligned} \tag{11}$$

Thus, this multiplicative weighted MPM (MWMPM) estimate of \mathbf{X} is equivalent to the MPM estimate of $\tilde{\mathbf{X}}$.

Note that in (Monaco and Madabhushi, 2011) we introduced another possible generalization of the MPM cost function. However, this generalization leads to a very different estimation strategy (see Appendix A). Also note that since the probability distribution of every MRF can be expressed using a Gibbs formulation (Besag, 1974), it is insightful to examine the Gibbs formulations of both $P(\mathbf{X} = \mathbf{x})$ and $P(\tilde{\mathbf{X}} = \mathbf{x})$. This examination is presented in Appendix B.

4. Implementations and Algorithms

In this section we demonstrate how MWMAP and MWMPM estimation can be performed using modifications of existing estimation strategies. As mentioned in Sections 3.1 and 3.2, the MWMAP and MWMPM estimates of a random field \mathbf{X} are equivalent to the MAP and MPM estimates of the random field $\tilde{\mathbf{X}}$, respectively. Consequently, we are free to realize these estimates using any existing MAP or MPM implementation. Though a variety of MAP estimation techniques exist (e.g. simulated annealing (Geman and Geman, 1984)), for MWMAP estimation we elect to employ iterated conditional modes (ICM) because of its popularity and simplicity. For MWMPM we select the Markov chain Monte Carlo (MCMC) method proposed by Marroquin *et al.* (Marroquin *et al.*, 1987).

4.1. Weighted Maximum a Posteriori (MWMAP) Estimation with Iterated Conditional Modes

ICM is predicated on the following reformulation of the *a posteriori* probability (Besag, 1986):

$$P(\mathbf{x}|\mathbf{y}) = P(x_s|\mathbf{x}_{-s}, \mathbf{y})P(\mathbf{x}_{-s}|\mathbf{y}). \tag{12}$$

Increasing $P(x_s|\mathbf{x}_{-s}, \mathbf{y})$ necessarily increases $P(\mathbf{x}|\mathbf{y})$. This suggests an optimization strategy that sequentially visits each site $s \in S$ and determines the label $x_s \in \Lambda$ that maximizes $P(x_s|\mathbf{x}_{-s}, \mathbf{y})$. The ICM algorithm is provided as Figure 1(a). Typically, the initial condition \mathbf{x}^0 is either a randomly selected element of Ω or the maximum likelihood estimate of $P(\mathbf{y}|\mathbf{x})$ (Pappas, 1992; Dubes *et al.*, 1990). ICM converges to a local maximum of $P(\mathbf{x}|\mathbf{y})$.

Since the MWMAP estimate of \mathbf{X} is equivalent to the MAP estimate of the $\tilde{\mathbf{X}}$, modifying ICM to perform MWMAP estimation only requires replacing $P(X_s = \omega|\mathbf{x}_{-s}^k, \mathbf{y})$ with $P(\tilde{X}_s = \omega|\tilde{\mathbf{X}}_{-s} = \mathbf{x}_{-s}^k, \mathbf{y})$ in step 6, and then recognizing (see Appendix B) that $P(\tilde{X}_s = \omega|\tilde{\mathbf{X}}_{-s} = \mathbf{x}_{-s}^k, \mathbf{y}) \propto \alpha(x_s)P(X_s = \omega|\mathbf{x}_{-s}^k, \mathbf{y})$. The resulting weighted ICM (WICM) algorithm is provided in Figure 1(b). Note that in practice, the computation of $P(x_s|\mathbf{x}_{-s}, \mathbf{y})$ is straightforward. Consider that $P(x_s|\mathbf{x}_{-s}, \mathbf{y})$ reduces to $P(x_s|\mathbf{x}_{\eta_s}, y_s)$ by consequence of the Markov property and the typical assumption that the observations \mathbf{Y} are conditionally independent

given their associated states, i.e. $P(\mathbf{y}|\mathbf{x}) = \prod_{s \in S} P(y_s|x_s)$. Furthermore, $P(x_s|\mathbf{x}_{\eta_s}, y_s) \propto P(y_s|x_s)P(x_s|\mathbf{x}_{\eta_s})$ by Bayes law.

4.2. Weighted Maximum Posterior Marginals (MWMPM) Using a Markov Chain Monte Carlo Simulation

MPM advocates selecting the estimate \mathbf{x} that maximizes the marginal probabilities $P(x_s|\mathbf{y})$ for all $s \in S$. To obtain these marginals Marroquin *et al.* proposed using the Gibbs sampler (Geman and Geman, 1984; Casella and George, 1992) or the Metropolis algorithm (Metropolis et al., 1953) to generate a Markov chain $(\mathbf{X}^0, \mathbf{X}^1, \mathbf{X}^2, \dots)$ with equilibrium distribution $P(\mathbf{x}|\mathbf{y})$, where \mathbf{X}^k is a random variable indicating the state of the chain at iteration k (see Figure 2(a)). Thus, the proportion of time the chain spends (after reaching equilibrium) in any state \mathbf{x} is given by $P(\mathbf{x}|\mathbf{y})$, i.e. each state \mathbf{x}^k represents a sample from the distribution $P(\mathbf{x}|\mathbf{y})$. The convergence to $P(\mathbf{x}|\mathbf{y})$ is independent of the starting conditions (Tierney, 1994); and consequently, \mathbf{x}^0 is typically selected at random from Ω . Determining the number of iterations l needed for the Markov chain to reach equilibrium is difficult, and depends upon the particular distribution $P(\mathbf{x}|\mathbf{y})$ and the initial conditions \mathbf{x}^0 . Usually l is selected empirically.

Since the proportion of time the chain spends in any state \mathbf{x} is given by $P(\mathbf{x}|\mathbf{y})$, the posterior marginal $P(x_s|\mathbf{y})$ can be estimated as follows:

$$P(X_s=\omega|\mathbf{y}) \approx \frac{1}{m-l} \sum_{k=l+1}^m \delta(x_s^k-\omega), \quad (13)$$

where $\omega \in \Lambda$ and $m-l$ is the number of iterations past equilibrium needed to generate an accurate estimate. The value for m , like l , is typically chosen empirically².

Since the MWMPM estimate of \mathbf{X} is equivalent to the MPM estimate of $\tilde{\mathbf{X}}$, modifying the previous MCMC method to explicitly perform MWMPM estimation only requires replacing $P(X_s=x_s^k|\mathbf{x}_{-s}^k, \mathbf{y})$ in step 6 of the Gibbs sampler (Figure 2(a)) with

$$P(\tilde{X}_s=x_s^k|\tilde{\mathbf{X}}_{-s}=\mathbf{x}_{-s}^k, \mathbf{y}) = \frac{\alpha(x_s^k)}{A_{\alpha s}} P(X_s=x_s^k|\mathbf{x}_{-s}^k, \mathbf{y}), \quad (14)$$

where $A_{\alpha s} = \sum_{\omega \in \Lambda} \alpha(\omega) P(X_s=\omega|\mathbf{x}_{-s}^k, \mathbf{y})$. We refer to this modified version of the Gibbs sampler as the weighted Gibbs sampler (see Figure 2(b)). The remainder of the estimation procedure is identical to that used for MPM, i.e. for all $s \in S$ we identify the $\omega \in \Lambda$ that maximizes the marginal distribution $P(\tilde{X}_s = \omega|\mathbf{y})$ obtained from (13).

5. Experimental Design

In this section, we begin by alternately incorporating the MWMAP and MWMPM estimation criteria into two MRF-based classification systems for detecting CaP: 1) a system for detecting CaP glands in HSs from radical prostatectomies and 2) a system for detecting CaP regions in T2-weighted 4 Telsa *ex vivo* prostate MRI. We then integrate these estimation criteria into a simple MRF-based classifier used to segment a two-dimensional synthetic brain MR image into regions containing cerebrospinal fluid (CSF), gray matter (GM), or white matter (WM). For each of these classification tasks, our goal is to

²Dubes and Jain (Dubes et al., 1990) refer to l and m as “magic” numbers.

demonstrate that by varying the class-Specific weights inherent in the MWMAP and MWMPM criteria, we can arbitrarily adjust their detection sensitivity/Specificity and generate ROC curves — or surfaces when the number of classes exceed two. To implement the MWMAP and MWMPM criteria, we employ the methods presented in Section 4, which we will refer to as MWMAP^{ICM} and MWMPM^{MC}. The superscripts ICM and MC (i.e. Monte Carlo) help describe the Specific approach, and hopefully, reemphasize that the estimation criteria and their Specific realizations should not be conflated. For convenience, the two estimation criteria and their associated implementations are listed in Table 2.

Before continuing, we should further clarify the difference between an estimation criterion and its associated implementation. In this paper, the MAP, MPM, MWMAP, and MWMPM estimation criteria are defined by equations (3), (5), (9), and (11), respectively. For a given estimation criterion, the “optimal” classification is the $\hat{\mathbf{x}}$ that minimizes that criterion. Solving for the optimal classification requires a Specific implementation. Because of complexity of MRFs, straightforward implementations that precisely minimize the estimation criteria do not exist; consequently, researchers have developed methods for yielding suboptimal classifications. For example, ICM (Besag, 1986), graph cuts (Boykov et al., 2001), and simulated annealing (Geman and Geman, 1984) are three approaches for approximating the MAP estimate, i.e. minimizing (3). That is, ICM, graph cuts, and simulated annealing are three possible implementations of the MAP estimation criterion. We should point out that each implementation is usually Specific to an estimation criteria; for example, ICM performs MAP estimation, and would not be used for MPM estimation.

5.1. Generating Receiver Operator Characteristic Curves and Surfaces

5.1.1. Binary Classes: ROC Curves—Both CaP detection systems are similar in the sense that they employ a MRF framework to classify their respective sites (i.e. glands or pixels) as either malignant ω_1 or benign ω_2 . Assuming the systems use one of our two estimation criteria, the classification results will depend upon the choice of weights $\alpha(\omega_1)$ and $\alpha(\omega_2)$. Since only the ratio of weights, and not their Specific values, is relevant, they can be represented using a single, more intuitive threshold

$$T = \frac{\alpha(\omega_2)}{\alpha(\omega_1) + \alpha(\omega_2)}, \quad (15)$$

where $T \in [0, 1]$. To see that applying the two weights is equivalent to thresholding (the Bayes factor (Kass and Raftery, 1995)), we need only rewrite step 6 in Figures 1(b) and 2(b) in terms of T (not shown).

To assess CaP detection performance, we define the following: true positives (TP) are those segmented objects (glands for HSs and pixels in the MRI) identified as cancerous by both the expert-provided ground-truth and the automated system; true negatives (TN) are those segmented objects identified as benign by both the truth and the automated system, false positives (FP) are those segmented objects identified as benign by the truth and malignant by the automated system; and false negatives (FN) are those segmented objects identified as malignant by the truth and benign by the automated system. The true positive rate (TPR) and false positive rate (FPR) are given by $TP/(TP+FN)$ and $FP/(TN+FP)$, respectively. Note that the TPR and FPR are synonymous with the sensitivity and one minus the Specificity, respectively. A ROC curve (Metz, 1978) is a plot of the TPR vs. FPR. For each estimation criterion, we can generate a ROC curve by varying T from zero to one, measuring the resulting TP/FP/FN/TN across all images, and then computing the TPR and FPR.

5.1.2. Ternary Classes: ROC Surfaces—The brain MR image, unlike the images for CaP detection, contain three classes: CSF (ω_1), GM (ω_2), and WM (ω_3). Segmentation performance depends upon the choice of weights $\alpha(\omega_1)$, $\alpha(\omega_2)$, and $\alpha(\omega_3)$. Since only relative ratios of weights are relevant, we can, without loss of generality, restrict the weights as follows:

$$\alpha(\omega_1) + \alpha(\omega_2) + \alpha(\omega_3) = 1 \text{ and } \alpha(\omega_1), \alpha(\omega_2), \alpha(\omega_3) \in [0, 1]. \quad (16)$$

Instead of the above restrictions, we could have employed two thresholds analogous to the T defined in (15); however, we found the representation in (16) to be more intuitive.

To assess segmentation performance, we define the class-Specific true positives TP_{ω_i} for class ω_i , $i \in \{1, 2, 3\}$ as the number of pixels labeled as ω_i by both the classifier and the ground-truth. The class-Specific true positive rate (TPR_{ω_i}) for class ω_i is $TP_{\omega_i} / \sum_i TP_{\omega_i}$. The triplet $(TPR_{\omega_1}, TPR_{\omega_2}, TPR_{\omega_3})$ — which is a function of the weights — establishes an operating point on the associated three-dimensional ROC surface. We populate this surface (with unique operating points) by appropriately varying the weights in (16).

5.2. Experiment 1: Detection of CaP Glands in Digitized Radical Prostatectomy Sections

The analysis of HSs plays a significant role in the diagnosis and treatment of CaP (Kumar et al., 2004). The most salient information in these HSs is derived from the morphology and architecture of the glandular structures (Gleason, 1966). Since complex tasks such as Gleason grading (Tabesh et al., 2007; Doyle et al., In Press) consider only the cancerous glands, an initial process capable of rapidly identifying these glands is highly desirable. Thus, we introduced an automated system for detecting cancerous glands in Hematoxylin and Eosin (H&E) stained tissue sections (Monaco et al., 2010). The primary goal of this algorithm is to eliminate regions of glands that are not likely to be cancerous, thereby reducing the computational load of further, more sophisticated analyses. Consequently, in a clinical setting the algorithm should operate at a high detection sensitivity, ensuring that very little CaP is discarded.

Figure 3(a) illustrates an H&E stained prostate histological (tissue) section. The superimposed black line delimits the spatial extent of CaP as determined by a pathologist. The numerous white regions are the gland lumens, i.e. cavities in the prostate through which fluid flows. Our system identifies CaP by leveraging two biological properties: 1) cancerous glands tend to be smaller in cancerous than benign regions and 2) malignant/benign glands tend to be proximate to other malignant/benign glands (Kumar et al., 2004). The basic algorithm proceeds as follows: Step 1) The glands (or, more precisely, the gland lumens) are identified and segmented. (Figure 3(b)–(c)). Step 2) Morphological features are extracted from the segmented boundaries. Currently, we consider only one feature: glandular area. Step 3) Using this feature and an MRF prior which encourages neighboring glands to share the same label, a Bayesian estimator classifies each gland as either malignant or benign (Figure 3(d)).

We now formally express this CaP detection problem using the MRF nomenclature established in Section 2.1. Let the set $S = \{1, 2, \dots, N\}$ reference the N segmented glands in a HS. Each site has an associated state $X_s \in \Lambda \equiv \{\omega_1, \omega_2\}$, where ω_1 and ω_2 indicate malignancy and benignity, respectively. The random variable $Y_s \in \mathbb{R}$ indicates the area of gland s . All feature Y_s are assumed conditionally independent and identically distributed (i.i.d.) given their corresponding states. The conditional distributions $P(Y_s | X_s)$ are modeled parametrically using a mixture of Gamma distributions (Monaco et al., 2010); this distribution is fit from training samples using maximum likelihood estimation. The tendency

for neighboring glands to share the same label is incorporated with a Markov prior $P(\mathbf{x})$ modeled using probabilistic pairwise Markov models (PPMMs) (Monaco et al., 2010) (see Appendix C); the PPMM is trained using maximum pseudo-likelihood estimation (MPLE) (Besag, 1986). Two glands are considered neighbors if the distance between their centroids is less than 0.9 mm.

We construct two ROC curves by performing classification using $MWMAPI^{ICM}$ and $MWMPM^{MC}$. Since both require that T be specified before running the relaxation procedure, we evaluate these configurations at 21 predetermined thresholds: $T \in \{0, 0.05, 0.1, \dots, 0.95, 1\}$. The TP, FP, TN, and FN are generated over 40 HSs from 20 patients using leave-one-out cross-validation. Ground-truth for each HS was delineated by an expert pathologist using either the physical specimen or its digitized image. All glands whose centroids fall within the truth are considered malignant; otherwise they are benign. The parameters used by $MWMPM^{MC}$ are as follows: $m=30$ and $l=10$.

5.3. Experiment 2: CaP Detection in MRI

Magnetic Resonance Imaging (MRI) has recently emerged as a promising modality for the non-invasive identification of CaP *in vivo* (Chelsky et al., 1993; Bloch et al., 2008). Because of its relatively high contrast and resolution it offers a possible means for guiding biopsies (Yu and Hricak, 2000), enhancing treatment (Yu and Hricak, 2000) (e.g. brachytherapy seed placement, high-focused ablation therapy), and improving screening for early detection. Motivated by these promising applications, we have developed automated systems for identifying CaP regions in MR images (Madabhushi et al., 2005; Tiwari et al., 2011; Viswanath et al., 2012; Monaco et al., 2009b).

For this experiment, we employ our system (Chappelow et al., 2008) for detecting CaP regions on T2-weighted 4 Tesla *ex vivo* prostate MR images. This detection occurs on a pixel-wise basis for each MR image. Figure 4(a) illustrates a typical MR image; the green overlay indicates the cancerous extent, determined by mapping pathologist-delineated CaP regions from a histological specimen onto a corresponding MRI section following the elastic registration (Chappelow et al., 2011) of the two modalities. The detection algorithm proceeds as follows: Step1) From each MRI section, we extract 21 gradient and statistical features at each pixel (Chappelow et al., 2008) (see Figures 4(b)–(c)). Step2) Using these features in combination with a Markov prior, which models the tendency for neighboring pixels to share the same class, each pixel is labeled as benign or malignant (Figure 4(d)).

This detection problem is recapitulated using the MRF terminology: Let the set $S = \{1, 2, \dots, N\}$ reference the N pixels in the MR image that reside within the prostate. Each site has an associated state $X_s \in \Lambda \equiv \{\omega_1, \omega_2\}$, where ω_1 and ω_2 indicate malignancy and benignity, respectively. The random vector $Y_s \in \mathbb{R}^D$ represents the $D = 21$ features associated with pixel s . All feature vectors Y_s are assumed conditionally independent and identically distributed (i.i.d.) given their corresponding states. Each multi-dimensional distribution $P(y_s|x_s)$ (one for each class) is modeled as the product of 21 one-dimensional histograms (i.e. the individual features are assumed independent). The tendency for neighboring pixels to share the same class is incorporated using a PPMM trained using MPLE. The neighborhood η_s of a pixel s is the typical 8-connected region.

We construct two ROC curves by performing classification using $MWMAPI^{ICM}$ and $MWMPM^{MC}$. Both $MWMAPI^{ICM}$ and $MWMPM^{MC}$ are evaluated at 21 thresholds: $T \in \{0, 0.05, 0.1, \dots, 0.95, 1\}$. The TP, FP, TN, and FN are generated over 15 (256×256) MRI slices from a single patient using leave-one-out cross-validation. As previously mentioned, ground-truth for each slice was determined by mapping pathologist-delineated CaP regions from a histological specimen onto a corresponding MRI section following registration

(Chappelow et al., 2011). All pixels lying within the pathologist-delineated CaP regions are considered malignant; otherwise they are benign. The parameters used by MWMPM^{MC} are as follows: $m = 50$ and $l = 30$.

5.4. Experiment 3: Segmentation of Brain MR Phantom

Experiments 1 and 2 both concern binary-class problems. We now present a task involving three-classes: segmenting brain MR images into regions of CSF, GM, and WM.

Specifically, we consider a single two-dimensional brain MR phantom from the BrainWeb Simulated Brain Database (Collins et al., 1998). This image, as obtained from the database, has 9 percent additive noise and 40 percent intensity non-uniformity. To increase the difficulty of the segmentation task, we inserted an additional 10 percent additive Gaussian noise, creating the image shown in Figure 5(a). Figure 5(b), also obtained from BrainWeb, depicts the ground-truth, which partitions the image into its three constituent regions: CSF (black), GM (gray), and WM (white).

We now describe the Specific classification procedure. Let the set $S = \{1, 2, \dots, N\}$ reference the N pixels in the MR image that reside within the brain. Each site has as associated state $X_s \in \Lambda \equiv \{\omega_1, \omega_2, \omega_3\}$, where ω_1 , ω_2 , and ω_3 indicate CSF, GM, and WM, respectively. The random vector $Y_s \in \mathbb{R}$ represents the MR intensity associated with pixel s . All random variables Y_s are assumed conditionally independent and identically distributed (i.i.d.) given their corresponding states. Each distribution $P(y_s|x_s)$ (one for each class) is modeled as a Gaussian density; the mean and standard deviation were determined using MLE. The tendency for neighboring pixels to share the same class is incorporated using a Potts MRF prior with $\beta=1$. The neighborhood η_s of a pixel s is the typical 8-connected region. Note, an example of the automated classification results are shown in Figure 5(c).

We construct two ROC surfaces (He et al., 2006) by performing classification using MWMAP^{ICM} and MWMPM^{MC} on the single (181×217) brain MR phantom. Both MWMAP^{ICM} and MWMPM^{MC} are evaluated for all combinations of weights

$\alpha(\omega_i) \in \{0, \frac{1}{100}, \frac{2}{100}, \dots, \frac{99}{100}, 1\}$ that satisfy (16). (This yields 4831 unique combinations.) For each combination, we determine the triplet $(\text{TPR}_{\omega_1}, \text{TPR}_{\omega_2}, \text{TPR}_{\omega_3})$. The parameters used by MWMPM^{MC} are $m=20$ and $l = 10$.

6. Results and Discussion

6.1. Experiment 1

Figure 6 illustrates the experimental results of the CaP gland detection system for the HSs. Figures 6(a) and 6(e) indicate the ROC curves when employing MWMAP^{ICM} and MWMPM^{MC}, respectively. The black dots indicate the performance at the 21 different T values. The connecting lines segments result from linearly interpolating between the points. Figures 6(b)–(d) and 6(f)–(h) provide qualitative examples of the final classification results for the estimation techniques at three different values of T . The green dots indicate the centroids of those glands labeled as malignant. The system performances at these T values are indicated with black circles in the corresponding ROC curves.

6.2. Experiment 2

Figure 7 presents analogous results for the CaP detection system for MR images. Figures 7(a) and 7(e) indicate the ROC curves using MWMAP^{ICM} and MWMPM^{MC}. The figures in Figures 7(b)–(d) and 7(f)–(h) depict the qualitative classification results for these estimation schemes at different values of T . The pixels classified as malignant are overlaid in red. The system performances at these T values are indicated with black circles in the corresponding ROC curves.

6.3. Experiment 3

Figure 8 illustrates the results of brain segmentation. Figures 8(a) and 8(e) indicate the ROC surfaces when employing $MWMAP^{ICM}$ and $MWMPM^{MC}$, respectively. Figures 8(b)–(d) and 8(f)–(h) provide qualitative examples of the final classification results for our two estimation techniques at three different combinations of weights.

6.4. Discussion

Notice that the ROC curves/surfaces produced by $MWMAP^{ICM}$ and $MWMPM^{MC}$ are similar. This is not unexpected since both implementations have related underlying cost functions. Specifically, they identify the MAP and MPM estimates of the same random variable $\tilde{\mathbf{X}}$ (see Section 3.1). It is important to point out that for the purposes of this paper the actual performances (e.g. the areas/volumes under the ROC curves/surfaces) of the classifiers are immaterial; the goal of this work is to demonstrate how the performance of any MRF-based classification system — that uses either MAP or MPM estimation — can be varied via the appropriate incorporation of multiplicative weights.

It is worth mentioning that our use of random initial conditions (\mathbf{x}^0) to begin each estimation technique is not optimal with respect to classification performance. We employed these conditions for two reasons: 1) to avoid any implication of a connection between the weights and the initial conditions and 2) to emphasize the importance of the estimation criteria (e.g. $MWMAP$) over any Specific implementation (e.g. $MWMAP^{ICM}$). Had we instead used the MLE of $P(\mathbf{y}|\mathbf{x})$, the resulting performance of both estimation schemes would have improved. With $MWMAP^{ICM}$, which converges to a local maximum of $P(\mathbf{x}|\mathbf{y})$, this is expected. However, even the Monte Carlo methods, whose convergence to $P(\mathbf{x}|\mathbf{y})$ is theoretically independent of the initial conditions, would benefit.

As is typical in ROC analysis, the curves/surfaces in Figures 6, 7, and 8 consist of a finite number of operating points. This discrete sampling is not a product of our novel weighting scheme, but results even when analyzing independent random variables for which likelihood ratios can be determined (He et al., 2006). However, producing ROC curves using our multiplicative weights — unlike the more familiar technique which employs likelihoods — requires rerunning the relaxation process (e.g. $WICM$) with every change in weights α . Thus, calculating each operating point can be a time-consuming process. (Note that using likelihoods to compute ROC curves is not straightforward with MRFs, and for MAP estimation it appears intractable.)

The need to perform multiple relaxations is a potential disadvantage of our proposed weighting scheme. However, this disadvantage only becomes problematic when the total time required to compute the necessary number of operating points is high. This computation time depends upon the total number of operating points — which is mostly a function of the number of classes — and the relaxation time — which is a function of the Specific classification system. Thus, it is difficult to predict the time needed to sufficiently sample ROC curves in absence of a definitive application. However, it is instructive to consider ROC construction for our three tasks.

Table 3 lists the total times needed to generate the ROC curves/surfaces depicted in Figures 6, 7, and 8. All times assume the use of single core of an Intel 2.5 GHz CPU. From these results, it becomes clear that both $MWMAP^{ICM}$ and $MWMPM^{MC}$ — at least for the three tasks presented in this paper — are computationally quite reasonable. However, extrapolating these results to predict time requirements for future applications should be done with care; as mentioned previously, total times will depend upon a host of factors such

as the number of sites, number of images, algorithm implementations, convergence times for WICM, and the number of iterations specified for the Weighted Gibbs sampler.

7. Concluding Remarks

Since MRFs typically contain large numbers of dependent random variables, estimating the state of the entire MRF is challenging, and requires sophisticated strategies. Currently these strategies weight each classification outcome equally, and consequently, provide no means for varying classifier performance. This is especially significant in medical image analysis, where certain errors (e.g. overlooking evidence of disease) are far more costly than others (e.g. further investigating an incorrect diagnosis of disease). Addressing this deficiency, we introduced MWMAP and MWMPM estimation, novel extensions of MAP and MPM estimation that allow certain classes to be favored more heavily than others. This creates a natural bias for Specific classes, and consequently a means for adjusting classifier performance. Additionally, we described how existing means for performing MAP and MPM estimation could be extended to obtain MWMAP and MWMPM estimates. To illustrate the value of our novel estimation criteria we incorporated them into two medically relevant MRF-based classification systems for detecting carcinoma of the prostate on 1) digitized HSs from radical prostatectomies and 2) T2-weighted 4 Tesla *ex vivo* prostate MRI. Furthermore, to underscore the extensibility of MWMAP and MWMPM to estimation tasks involving more than two classes, we also incorporated these estimation criteria into a MRF-based classifier used to segment synthetic brain MR images. In the context of these three tasks, we demonstrated how MWMAP and MWMPM estimation schemes could arbitrarily vary the cancer detection sensitivity of these systems, yielding receiver operator characteristic curves and surfaces.

Before concluding, it is worthwhile to briefly consider another technique that could be used to vary the performance of MRF-based classifiers: fuzzy MRFs (Ruan et al., 2002; Salzenstein and Collet, 2006). In theory, thresholds could be applied to each site's fuzzy membership values, yielding different classifications. However, fuzzy membership was intended to indicate the degree to which a single site belongs to each of the possible classes (e.g. to account for partial volume effects), and not to reflect the probability of belonging to a Specific class. Thus, constructing ROC curves in this manner appears heuristic.

Acknowledgments

This work was made possible via grants from the Wallace H. Coulter Foundation, National Cancer Institute (Grant Nos. R01CA136535-01, R01CA14077201, and R03CA143991-01), and the Cancer Institute of New Jersey. Both authors are major shareholders of IbRiS Inc. We would like to thank Satish Viswanath for his help with the prostate MRI application.

References

- Awate, S.; Tasdizen, T.; Whitaker, R. Computer Vision ECCV. 2006. Unsupervised texture segmentation with nonparametric neighborhood statistics; p. 494-507.
- Besag J. Spatial interaction and the statistical analysis of lattice systems. Journal of the Royal Statistical Society Series B (Methodological). 1974; 36(2):192–236. URL <http://www.jstor.org/stable/2984812>.
- Besag J. On the statistical analysis of dirty pictures. Journal of the Royal Statistical Society Series B (Methodological). 1986; 48(3):259–302. URL <http://www.jstor.org/stable/2345426>.
- Bloch BN, Lenkinski RE, Rofsky NM. The role of magnetic resonance imaging (mri) in prostate cancer imaging and staging at 1.5 and 3 tesla: the beth israel deaconess medical center (bidmc) approach. Cancer Biomark. 2008; 4 (4–5):251–262. [PubMed: 18957714]

- Bouman CA, Shapiro M. A multiscale random field model for bayesian image segmentation. *IEEE Transactions on Image Processing*. Mar; 1994 3 (2):162–177. [PubMed: 18291917]
- Boykov Y, Veksler O, Zabih R. Fast approximate energy minimization via graph cuts. *Transactions on Pattern Analysis and Machine Intelligence*. Nov; 2001 23 (11):1222–1239.
- Casella G, George E. Explaining the gibbs sampler. *The American Statistician*. 1992; 46(3):167–174. URL <http://www.jstor.org/stable/2685208>.
- Chappelow J, Bloch N, Rofsky N, Genega E, Lenkinski R, DeWolf W, Madabhushi A. Elastic registration of multimodal prostate mri and histology via multiattribute combined mutual information. *Medical Physics*. 2011; 38 (4):2005–2018. [PubMed: 21626933]
- Chappelow, J.; Viswanath, S.; Monaco, J.; Rosen, M.; Tomaszewski, J.; Feldman, M.; Madabhushi, A. Improving supervised classification accuracy using non-rigid multimodal image registration: detecting prostate cancer. Vol. 6915. *SPIE*; Mar. 2008 p. 69150V URL <http://link.aip.org/link/?PSI/6915/69150V/1>
- Chelsky MJ, Schnall MD, Seidmon EJ, Pollack HM. Use of endorectal surface coil magnetic resonance imaging for local staging of prostate cancer. *J Urol*. Aug; 1993 150 (2 Pt 1):391–395. [PubMed: 8326561]
- Collins DL, Zijdenbos AP, Kollokian V, Sled JG, Kabani NJ, Holmes CJ, Evans AC. Design and construction of a realistic digital brain phantom. *IEEE Trans Med Imaging*. Jun; 1998 17(3):463–468. URL <http://dx.doi.org/10.1109/42.712135>. [PubMed: 9735909]
- Comer ML, Delp EJ. Segmentation of textured images using a multiresolution gaussian autoregressive model. *IEEE Transaction on Image Processing*. Mar; 1999 8 (3):408–420.
- Doyle S, Monaco J, Tomaszewski J, Feldman M, Madabhushi A. Active learning of minority classes: application to prostate histopathology annotation. *BMC Bioinformatics*. In Press.
- Dubes, R.; Jain, A.; Nadabar, S.; Chen, C. Mrf model-based algorithms for image segmentation. *Proceedings of the 10th International Conference on Pattern Recognition*; 1990. p. 808-814.
- Duda, R.; Hart, P.; Stork, D. *Pattern Classification*. John Wiley & Sons; 2001.
- Farag A, El-Baz A, Gimel'farb G. Precise segmentation of multimodal images. *IEEE Transactions on Image Processing*. Apr; 2006 15 (4):952–968. [PubMed: 16579381]
- Figueiredo MAT, Leitao JMN. Unsupervised image restoration and edge location using compound gauss-markov random fields and the mdl principle. *IEEE Transactions on Image Processing*. Aug; 1997 6 (8):1089–1102. [PubMed: 18282999]
- Geman, D. *Lecture Notes in Mathematics*. Vol. 1427. Springer-Verlag; 1991. Random fields and inverse problems in imaging; p. 113-193.
- Geman S, Geman D. Stochastic relaxation, gibbs distribution, and the bayesian restoration of images. *IEEE Transactions on Pattern Recognition and Machine Intelligence*. 1984; 6:721–741.
- Gleason D. Classification of prostatic carcinomas. *Cancer chemotherapy reports*. 1966; 50:125–128.
- He X, Metz C, Tsui B, Links J, Frey E. Three-class roc analysis-a decision theoretic approach under the ideal observer framework. *Medical Imaging, IEEE Transactions on*. May; 2006 25 (5):571–581.
- Kass R, Raftery A. Bayes factors. *Journal of the American Statistical Association*. 1995; 90(430):773–795. URL <http://www.jstor.org/stable/2291091>.
- Kumar, V.; Abbas, A.; Fausto, N. *Robbins and Cotran Pathologic Basis of Disease*. Saunders; 2004.
- Liu X, Langer D, Haider M, Yang Y, Wernick M, Yetik I. Prostate cancer segmentation with simultaneous estimation of markov random field parameters and class. *IEEE Transactions on Medical Imaging*. Jun; 2009 28 (6):906–915. [PubMed: 19164079]
- Madabhushi A, Feldman MD, Metaxas DN, Tomaszewski J, Chute D. Automated detection of prostatic adenocarcinoma from high-resolution ex vivo mri. *IEEE Trans Med Imaging*. Dec; 2005 24(12):1611–1625. URL <http://dx.doi.org/10.1109/TMI.2005.859208>. [PubMed: 16350920]
- Marroquin J, Mitter S, Poggio T. Probabilistic solution of ill-posed problems in computational vision. *Journal of the American Statistical Association*. 1987; 82(397):76–89. URL <http://www.jstor.org/stable/2289127>.

- Marroquin JL, Vemuri BC, Botello S, Calderon E, Fernandez-Bouzas A. An accurate and efficient bayesian method for automatic segmentation of brain mri. *IEEE Transactions on Medical Imaging*. Aug; 2002 21 (8):934–945. [PubMed: 12472266]
- Metropolis N, Rosenbluth AW, Rosenbluth MN, Teller AH, Teller E. Equation of state calculations by fast computing machines. *J Chem Phys*. Jun.1953 21:1087–1092.
- Metz C. Basic principles of roc analysis. *Seminars in Nuclear Medicine*. 1978; 8:283–298. [PubMed: 112681]
- Monaco J, Madabhushi A. Weighted maximum posterior marginals for random fields using an ensemble of conditional densities from multiple markov chain monte carlo simulations. *Medical Imaging, IEEE Transactions on PP*. 2011; (99):1.
- Monaco J, Tomaszewski J, Feldman M, Moradi M, Mousavi P, Boag A, Davidson C, Abolmaesumi P, Madabhushi A. Probabilistic pairwise markov models: Application to prostate cancer detection. *SPIE Medical Imaging*. 2009a:7260.
- Monaco, J.; Viswanath, S.; Madabhushi, A. Weighted iterated conditional modes for random fields: Application to prostate cancer detection. *Workshop on Probabilistic Models for Medical Image Analysis (in conjunction with MICCAI)*; 2009b.
- Monaco JP, Tomaszewski J, Feldman M, Hagemann I, Moradi M, Mousavi P, Boag A, Davidson C, Abolmaesumi P, Madabhushi A. High-throughput detection of prostate cancer in histological sections using probabilistic pairwise markov models. *Medical Image Analysis*. 2010; 14 (4):617–629. [PubMed: 20493759]
- Paget R, Longstaff ID. Texture synthesis via a noncausal nonparametric multiscale markov random field. *IEEE Transactions on Image Processing*. Jun; 1998 7 (6):925–931. [PubMed: 18276310]
- Pappas TN. An adaptive clustering algorithm for image segmentation. *IEEE Transaction on Signal Processing*. Apr; 1992 40 (4):901–914.
- Ruan S, Moretti B, Fadili J, Bloyet D. Fuzzy markovian segmentation in application of magnetic resonance images. *Computer Vision and Image Understanding*. 2002; 85(1):54–69. URL <http://www.sciencedirect.com/science/article/B6WCX-46W18DX-3/2/d3eb7cbedcdba3fb2914344f57af128c>.
- Salzenstein F, Collet C. Fuzzy markov random fields versus chains for multispectral image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 2006; 28:1753–1767. [PubMed: 17063681]
- Scherrer B, Forbes F, Garbay C, Dojat M. Distributed local mrf models for tissue and structure brain segmentation. *Medical Imaging, IEEE Transactions on*. Aug; 2009 28 (8):1278–1295.
- Szeliski R, Zabih R, Scharstein D, Veksler O, Kolmogorov V, Agarwala A, Tappen M, Rother C. A comparative study of energy minimization methods for markov random fields with smoothness-based priors. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. Jun; 2008 30 (6): 1068–1080. [PubMed: 18421111]
- Tabesh A, Teverovskiy M, Pang HY, Kumar VP, Verbel D, Kotsianti A, Saidi O. Multifeature prostate cancer diagnosis and gleason grading of histological images. *IEEE Transactions on Medical Imaging*. Oct; 2007 26 (10):1366–1378. [PubMed: 17948727]
- Tierney L. Markov chains for exploring posterior distributions. *The Annals of Statistics*. 1994; 22(4): 1701–1728. URL <http://www.jstor.org/stable/2242477>.
- Tiwari, P.; Viswanath, S.; Kurhanewicz, J.; Sridhar, A.; Madabhushi, A. Multimodal wavelet embedding representation for data combination (maweric): integrating magnetic resonance imaging and spectroscopy for prostate cancer detection. *NMR Biomed*. Sep. 2011 URL <http://dx.doi.org/10.1002/nbm.1777>
- Viswanath, SE.; Bloch, NB.; Chappelow, JC.; Toth, R.; Rofsky, NM.; Genega, EM.; Lenkinski, RE.; Madabhushi, A. Central gland and peripheral zone prostate tumors have significantly different quantitative imaging signatures on 3 tesla endorectal, in vivo t2-weighted mr imagery. *J Magn Reson Imaging*. Feb. 2012 URL <http://dx.doi.org/10.1002/jmri.23618>
- Yedidia J, Freeman W, Weiss Y. Generalized belief propagation. *Advances in Neural Information Processing Systems*. 2000:689–695.
- Yu KK, Hricak H. Imaging prostate cancer. *Radiol Clin North Am*. 2000; 38(1):59–85. viii. [PubMed: 10664667]

Zalesny, A.; Gool, LV. A compact model for viewpoint dependent texture synthesis. SMILE Workshop; Springer-Verlag, London, UK. 2001. p. 124-143.

Appendix

A. Alternative Generalization of Maximum Posterior Marginals

In Section 3.2 we generalized the MPM cost function in (4) by incorporating multiplicative weights, yielding Equation (10). However, alternative generalizations are possible. In (Monaco and Madabhushi, 2011) we incorporated additive weights to produce the following cost function:

$$C_{\text{WMPM}}(\mathbf{x}, \widehat{\mathbf{x}}) = \sum_{s \in \mathcal{S}} \alpha(x_s) [1 - \delta(x_s - \widehat{x}_s)], \quad (17)$$

which leads to a Bayesian risk given by

$$\begin{aligned} R_{\text{WMPM}}(\mathbf{X}[\widehat{\mathbf{x}}, \mathbf{y}]) &= \sum_{\mathbf{x} \in \Omega} \left(\sum_{s \in \mathcal{S}} \alpha(x_s) [1 - \delta(x_s - \widehat{x}_s)] \right) P(\mathbf{x}|\mathbf{y}) \\ &= \sum_{s \in \mathcal{S}} \sum_{\mathbf{x} \in \Omega} \alpha(x_s) P(\mathbf{x}|\mathbf{y}) - \sum_{s \in \mathcal{S}} \alpha(\widehat{x}_s) P(\widehat{x}_s|\mathbf{y}). \end{aligned} \quad (18)$$

Minimizing this risk function entails maximizing each of the weighted posterior marginals $\alpha(\widehat{x}_s)P(\widehat{x}_s|\mathbf{y})$, and thus — like MPM (and MWMPM) — requires estimates of these marginals. Unfortunately, the MCMC method (Marroquin et al., 1987) commonly used to estimate the posterior marginals — though sufficient for MPM (and MWMPM) estimation — is inadequate for this weighted extension of MPM. Consequently, in (Monaco and Madabhushi, 2011) we introduced a more appropriate estimation strategy.

B. Gibbs Formulation

The connection between the Markov property and the joint probability density function P of \mathbf{X} is revealed by the Hammersley-Clifford (Gibbs-Markov equivalence) theorem (Besag, 1974). This theorem states that a random field (G, Ω, P) with $P(\mathbf{x}) > 0$ for all $\mathbf{x} \in \Omega$ satisfies the Markov property if, and only if, it can be expressed as a Gibbs distribution:

$$P(\mathbf{x}) = \frac{1}{Z} \exp \left\{ \sum_{c \in \mathcal{C}} V_c(\mathbf{x}) \right\}, \quad (19)$$

where $Z = \sum_{\mathbf{x} \in \Omega} \exp\{\sum_{c \in \mathcal{C}} V_c(\mathbf{x})\}$ is the normalizing constant and V_c are functions, called clique potentials, that depend only on those x_s such that $s \in c$. A clique c is any subset of \mathcal{S} which constitutes a fully connected subgraph of G ; the set \mathcal{C} contains all possible cliques. Note that typically $|\Omega| = |\Lambda|^N$ is too large to directly evaluate Z . The following reveals the forms of the local conditional probability density functions:

$$P(x_s | \mathbf{x}_{\eta_s}) = \frac{1}{Z_s} \exp \left\{ \sum_{c \in \mathcal{C}_s} V_c(\mathbf{x}) \right\}, \quad (20)$$

where \mathcal{C}_s represents $\{c \in \mathcal{C} : s \in c\}$ and $Z_s = \sum_{\mathbf{x}_{\eta_s} \in \Lambda^{\eta_s}} \exp\{\sum_{c \in \mathcal{C}_s} V_c(\mathbf{x})\}$. For proofs of Markov formulations and theorems, see Geman (Geman, 1991).

It is now useful to consider the Gibbs formulation of the weighted probability function in (7):

$$\begin{aligned} P(\tilde{\mathbf{X}}=\mathbf{x}) &= \frac{1}{Z_\alpha} \alpha(\mathbf{x}) \exp \left\{ \sum_{c \in \mathcal{C}} V_c(\mathbf{x}) \right\} \\ &= \frac{1}{Z_\alpha} \exp \left\{ \sum_{s \in \mathcal{S}} \ln \alpha(x_s) + \sum_{c \in \mathcal{C}} V_c(\mathbf{x}) \right\} \quad (21) \\ &= \frac{1}{Z_\alpha} \exp \left\{ \sum_{c \in \mathcal{C}} \tilde{V}_c(\mathbf{x}) \right\}, \end{aligned}$$

where $Z_\alpha = \sum_{\mathbf{x} \in \Omega} \alpha(\mathbf{x}) \exp \{ \sum_{c \in \mathcal{C}} V_c(\mathbf{x}) \}$ is the normalizing constant and $\tilde{V}_c(\mathbf{x})$ is defined as follows: if $c = \{s\}$, $s \in \mathcal{S}$ then $\tilde{V}_c(\mathbf{x}) = V_c(\mathbf{x}) + \ln \alpha(x_s)$, otherwise $\tilde{V}_c(\mathbf{x}) = V_c(\mathbf{x})$. Thus, the weighting method proposed in this paper manifests as an increase of $\ln \alpha(x_s)$ in each single element clique potential. The forms of the local conditional probability density functions are as follows:

$$\begin{aligned} P(\tilde{X}_s = x_s | \tilde{\mathbf{X}}_{\eta_s} = \mathbf{x}_{\eta_s}) &= \frac{1}{Z_{\alpha s}} \exp \left\{ \sum_{c \in \mathcal{C}_s} \tilde{V}_c(\mathbf{x}) \right\} \\ &= \frac{1}{Z_{\alpha s}} \exp \left\{ \ln \alpha(x_s) + \sum_{c \in \mathcal{C}_s} V_c(\mathbf{x}) \right\} \quad (22) \\ &= \frac{1}{Z_{\alpha s}} \alpha(x_s) \exp \left\{ \sum_{c \in \mathcal{C}_s} V_c(\mathbf{x}) \right\} \\ &= \frac{Z_s}{Z_{\alpha s}} \alpha(x_s) P(x_s | \mathbf{x}_{\eta_s}), \end{aligned}$$

where $Z_{\alpha s} = \sum_{x_s \in \Lambda} \alpha(x_s) \exp \{ \sum_{c \in \mathcal{C}_s} V_c(\mathbf{x}) \}$.

C. Probabilistic Pairwise Markov Models

Before discussing probabilistic pairwise Markov models (PPMMs), we must first introduce additional notation. As discussed previously, $P(\cdot)$ indicates the probability of event $\{\cdot\}$. For instance, $P(X_s = x_s)$ and $P(\mathbf{X}=\mathbf{x})$ signify the probabilities of the events $\{X_s = x_s\}$ and $\{\mathbf{X}=\mathbf{x}\}$. Note that we simplified such notations in the paper — when it did not cause ambiguity — by omitting the random variable, e.g. $P(\mathbf{x}) \equiv P(\mathbf{X}=\mathbf{x})$. We now introduce $p(\cdot)$, which indicates a generic (discrete) probability function; for example, p_u might be a uniform distribution. The notations $P(\cdot)$ and $p(\cdot)$ are useful in differentiating $P(x_s)$ which indicates the probability that $\{X_s = x_s\}$ from $p_u(x_s)$ which refers to the probability that a uniform random variable assumes the value x_s .

Continuing, in place of potential functions (i.e. a Gibbs formulation), PPMMs (Monaco et al., 2010) formulate the local conditional probability density functions (LCPDFs) $P(x_s | \mathbf{x}_{\eta_s})$ of an MRF in terms of pairwise density functions, each of which models the interaction between two neighboring sites. This formulation facilitates the creation of relatively sophisticated LCPDFs (and hence priors), increasing our ability to model complex processes. Within the context of our CaP detection system, we previously demonstrated the superiority of PPMMs over the prevalent Potts model (Monaco et al., 2010). The PPMM formulation of the LCPDFs is as follows:

$$P(x_s | \mathbf{x}_{\eta_s}) = \frac{1}{Z_s} p_0(x_s) \prod_{r \in \eta_s} p_{1|0}(x_r | x_s), \quad (23)$$

where the normalizing constant Z_s ensures summation to one, p_0 is the probability density function (PDF) describing the stationary site s , and $p_{1|0}$ represents the conditional PDF

describing the pairwise relationship between site s and its neighboring site r . The numbers 0 and 1 replace the letters s and r to indicate that the probabilities are identical across all sites, i.e. the MRF is stationary. Furthermore, p_0 and $p_{1|0}$ are related in the sense that they are a marginal and conditional distribution of the joint distribution $p_{0,1}$, i.e. $p_{0,1}(x_s, x_r) = p_0(x_s)p_{1|0}(x_r, x_s)$. We are free to choose any forms for p_0 and $p_{1|0}$, under the caveat that $p_{0,1}$ be symmetric to ensure stationarity. Please see (Monaco et al., 2010) for further details.

\$watermark-text

\$watermark-text

\$watermark-text

Highlights

We introduce estimation criteria for adjusting the performance of MRF-based classifiers.

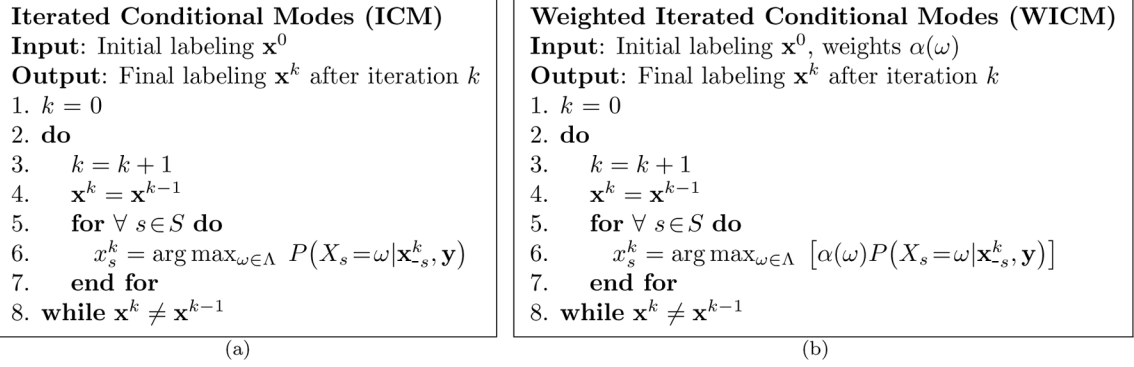
We implement these criteria adapting existing MRF algorithms (e.g. iterated conditional modes).

We integrate the implementations into three classification systems for medical image analysis.

\$watermark-text

\$watermark-text

\$watermark-text

**Figure 1.**

Algorithms for (a) iterated conditional modes and (b) weighted iterated conditional modes. Both algorithms are deterministic relaxation schemes that converge to a local maximum of $P(\mathbf{x}|\mathbf{y})$ and $\mathbf{a}(\mathbf{x})P(\mathbf{x}|\mathbf{y})$, respectively.

Gibbs Sampler**Input:** Initial labeling \mathbf{x}^0 **Output:** Final labeling \mathbf{x}^k after iteration k

1. $k = 0$
2. **do**
3. $k = k + 1$
4. $\mathbf{x}^k = \mathbf{x}^{k-1}$
5. **for** $\forall s \in S$ **do**
6. sample x_s^k from $P(X_s = x_s^k | \mathbf{x}_{-s}^k, \mathbf{y})$
7. **end for**
8. **while** $k < \infty$

(a)

Weighted Gibbs Sampler**Input:** Initial labeling \mathbf{x}^0 , weights $\alpha(\omega)$ **Output:** Final labeling \mathbf{x}^k after iteration k

1. $k = 0$
2. **do**
3. $k = k + 1$
4. $\mathbf{x}^k = \mathbf{x}^{k-1}$
5. **for** $\forall s \in S$ **do**
6. sample x_s^k from $\frac{\alpha(x_s^k)}{A_s} P(X_s = x_s^k | \mathbf{x}_{-s}^k, \mathbf{y})$
7. **end for**
8. **while** $k < \infty$

(b)

Figure 2.

Algorithms for the Gibbs sampler and the weighted Gibbs sampler. These two Monte Carlo algorithms generate Markov chains with equilibrium distributions $P(\mathbf{x}|\mathbf{y})$ and $P(\tilde{\mathbf{X}}=\mathbf{x}|\mathbf{y})$, respectively.

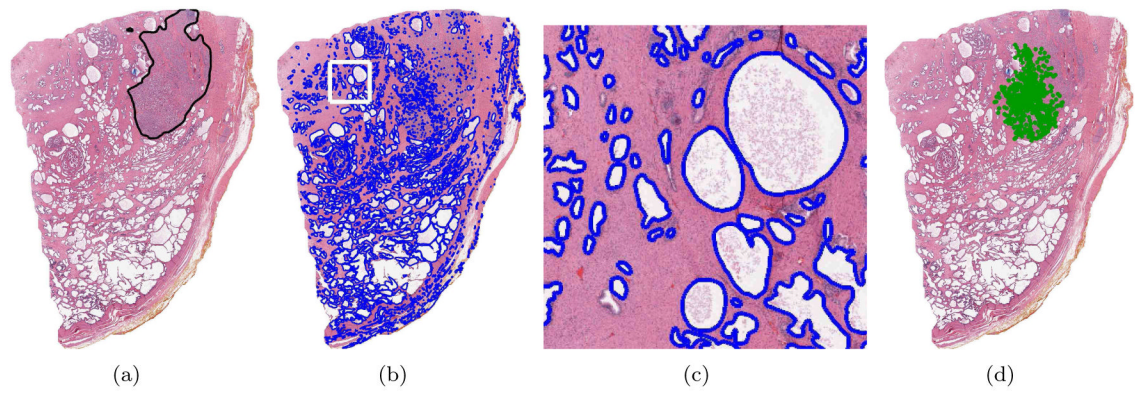


Figure 3.

(a) H&E stained prostate histology section; black ink mark indicates CaP extent as delineated by a pathologist. (b) Gland segmentation boundaries. (c) Magnified view of white box in (b). (d) Green dots indicate the centroids of those glands labeled as malignant.

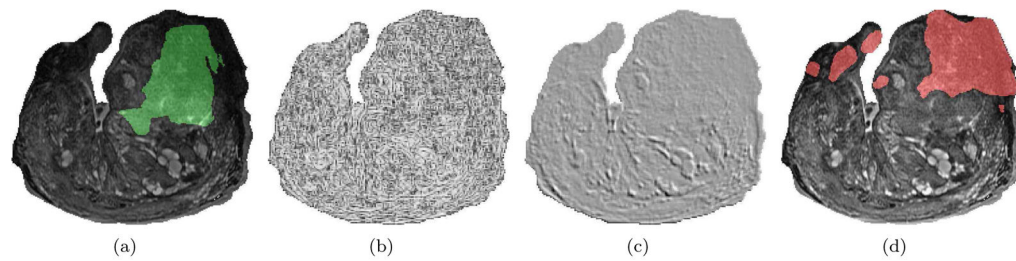


Figure 4.

(a) T2-weighted 4 Tesla *ex vivo* MRI of an excised prostate gland with cancerous region (overlayed in green) determined by mapping pathologist-delineated CaP regions from an associated histological specimen onto the MR image after registering the two modalities. (b), (c) Images illustrating two of the 21 gradient and statistical features. (d) Result of automated CaP detection (malignant pixels in red).

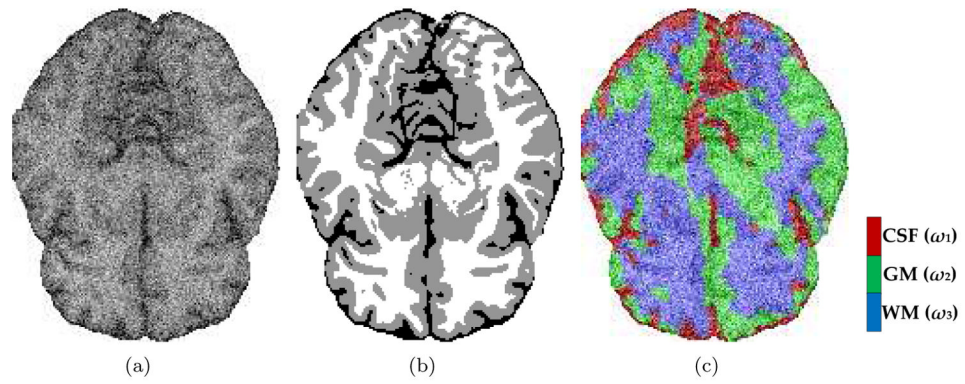


Figure 5.

(a) Two-dimensional simulated brain MR image (Collins et al., 1998) with 19 percent additive noise and 40 percent intensity non-uniformity. (b) Ground-truth image with ideal segmentation of CSF (black), GM (gray), and WM (white). (c) Automated segmentation results indicating CSF (red), GM (green), and WM (blue).

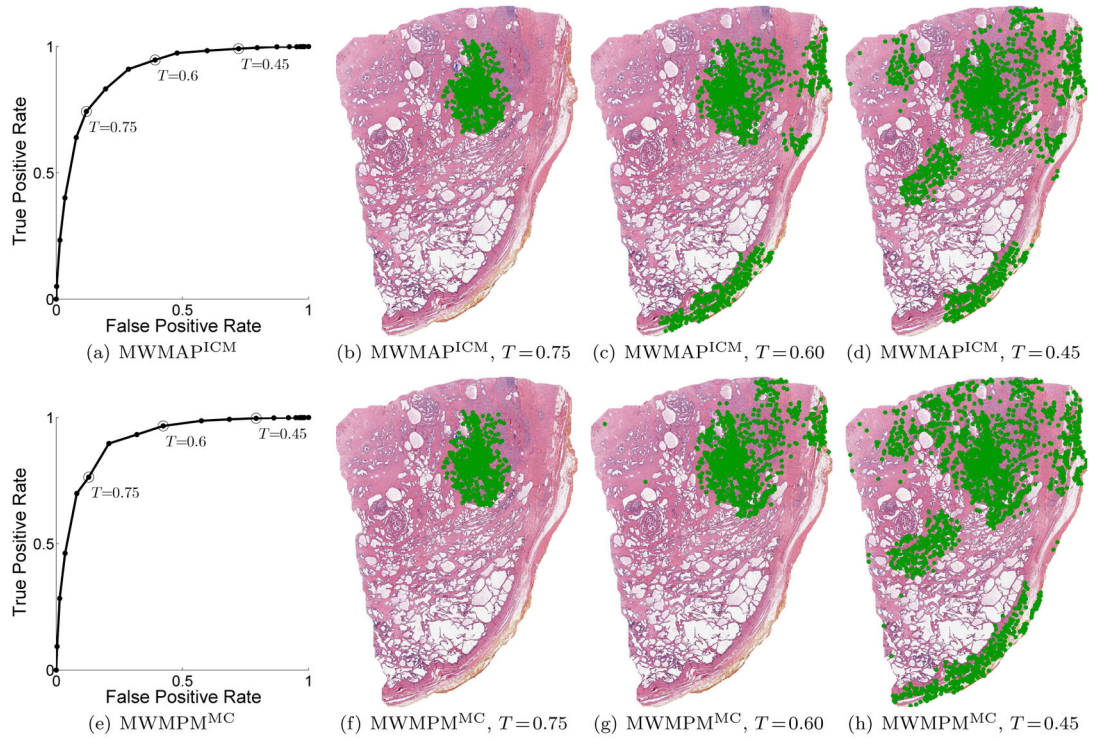


Figure 6.

(a), (e) ROC curves of CaP detection system on HSs using MWMAP^{ICM} and MWMPM^{MC}. The black dots in (a) and (e) indicate the performance at $T \in \{0, 0.05, 0.1, \dots, 0.95, 1\}$. (b)–(d) Centroids of the glands labeled as malignant (green dots) using MWMAP^{ICM} for $T \in \{0.75, 0.6, 0.45\}$. System performances at these T values are indicated by the hollow black circles in (a). (f)–(h) Centroids of the glands labeled as malignant using MWMPM^{MC} for $T \in \{0.75, 0.6, 0.45\}$. Corresponding system performances at these T values are indicated by the hollow black circles in (e).

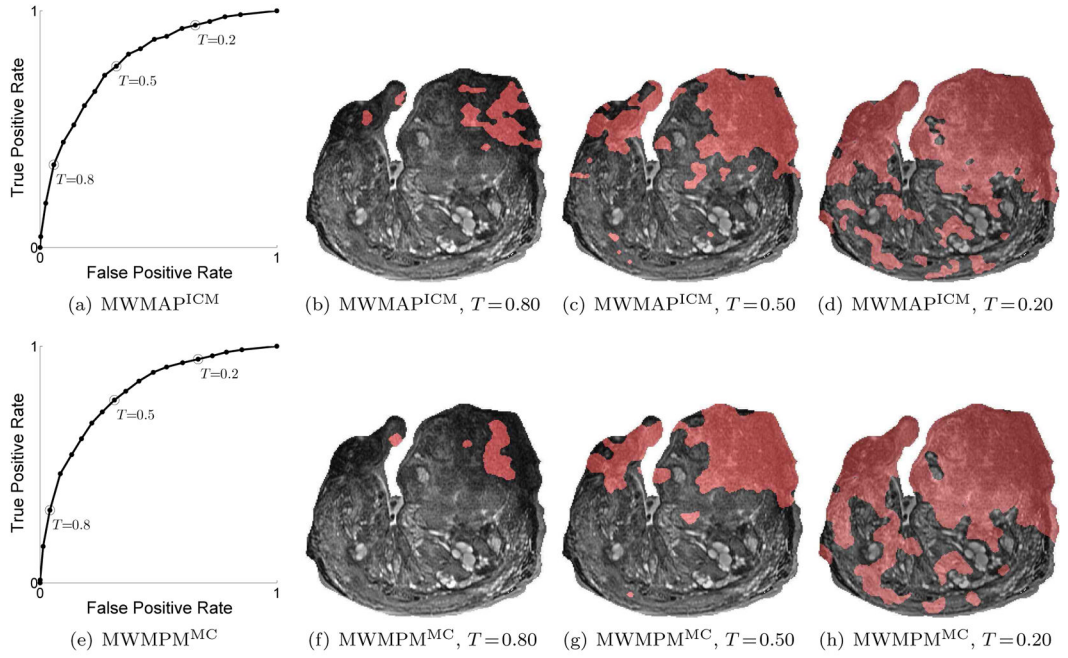


Figure 7. (a), (e) ROC curves of MRI CaP detection system using MWMAP^{ICM} and MWMPM^{MC}. The black dots in (a) and (e) indicate the performance for $T \in \{0, 0.05, 0.1, \dots, 0.95, 1\}$. (b)–(d) Pixels labeled as malignant (overlaid in red) using MWMAP^{ICM} for $T \in \{0.8, 0.5, 0.2\}$. The system performances at these T values are indicated by the hollow black circles in (a). (f)–(h) Pixels labeled as malignant using MWMPM^{MC} for $T \in \{0.8, 0.5, 0.2\}$. The system performances for these values of T are indicated by the hollow black circles in (e).

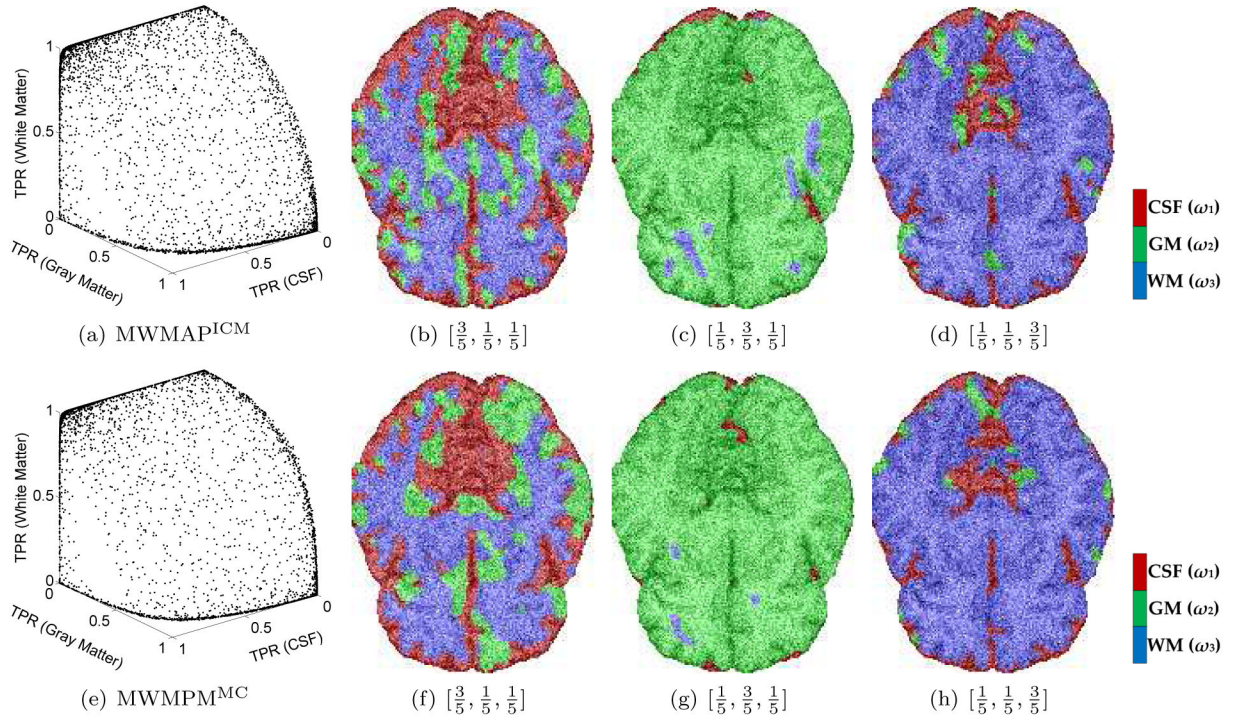


Figure 8.

(a), (e) ROC surfaces depicting brain MRI segmentation performance using MWMAP^{ICM} and MWMPM^{MC}. The black dots in (a) and (e) indicate the performance for all $\alpha(\omega_1), \alpha(\omega_2), \alpha(\omega_3) \in \{0, \frac{1}{100}, \frac{2}{100}, \dots, \frac{99}{100}, 1\}$ that satisfy (16). (b)–(d) Segmentations of MR image in Figure 5(a) using MWMAP^{ICM} for different combinations of weights $[\alpha(\omega_1), \alpha(\omega_2), \alpha(\omega_3)]$. Red, green, and blue colors indicate CSF, GM, and WM, respectively. The black dots in (f)–(h) indicate the performance for all $\alpha(\omega_1), \alpha(\omega_2), \alpha(\omega_3) \in \{0, \frac{1}{100}, \frac{2}{100}, \dots, \frac{99}{100}, 1\}$ that satisfy (16). (b)–(d) Segmentations of MR image in Figure 5(a) using MWMPM^{MC} for different combinations of weights $[\alpha(\omega_1), \alpha(\omega_2), \alpha(\omega_3)]$.

Table 1

List of notation and symbols.

Symbol	Description
S	Set referencing N sites
Λ	Range of X_s and x_s ; $\Lambda \equiv \{\omega_1, \omega_2, \dots, \omega_L\}$
$X_s \in \Lambda$	Random variable indicating state at site s
$x_s \in \Lambda$	Instance of X_s
D	Number of features
$Y_s \in \mathbb{R}^D$	Random variable indicating feature vector at site s
$y_s \in \mathbb{R}^D$	Instance of Y_s
$\hat{\mathbf{x}} \in \Omega$	Estimate of \mathbf{X}
$C(\mathbf{x}, \hat{\mathbf{x}})$	Cost of choosing $\hat{\mathbf{x}}$ when the true labels are \mathbf{x}
$R(\mathbf{X} \hat{\mathbf{x}}, \mathbf{y})$	Conditional risk: $R(\mathbf{X} \hat{\mathbf{x}}, \mathbf{y}) = E[C(\mathbf{X}, \hat{\mathbf{x}}) \mathbf{y}]$
$\mathbf{X} \in \Omega$	Collection of all X_s ; $\mathbf{X}=(X_1, X_2, \dots, X_N)$
$\mathbf{x} \in \Omega$	Instance of \mathbf{X} : $\mathbf{x}=(x_1, x_2, \dots, x_N)$
Ω	Range of \mathbf{X} and \mathbf{x} : $\Omega=\Lambda^N$
$\mathbf{Y} \in \mathbb{R}^{D \times N}$	Collection of all Y_s ; $\mathbf{Y}=(Y_1, Y_2, \dots, Y_N)$
$\mathbf{y} \in \mathbb{R}^{D \times N}$	Instance of \mathbf{Y} : $\mathbf{y}=(y_1, y_2, \dots, y_N)$
η_s	Set of sites that neighbor $s \in S$
\mathbf{x}_{-s}	$\mathbf{x}_{-s}=(x_1, \dots, x_{s-1}, x_{s+1}, \dots, x_N)$
\mathbf{x}_{η_s}	$\mathbf{x}_{\eta_s} = x_{\eta_s(1)}, \dots, x_{\eta_s(\eta_s)}$
$a(\cdot)$	Weighting function $a: \Lambda \rightarrow \mathbb{R}_0^+$
$\mathbf{a}(\mathbf{x})$	$\mathbf{a}(\mathbf{x})=\prod_{s \in S} a(x_s)$

Table 2

List of weighted estimation criteria and their associated implementations.

Estimation Criterion	Implementation
MWMAP	MWMAP ^{1CM}
MWMPM	MWMPM ^{1MC}

\$watermark-text

\$watermark-text

\$watermark-text

Table 3

Time required to populate ROC curves/surfaces.

Experiment (#)	Implementation	
	MWMAPI ^{ICM}	MWMPM ^{MC}
Prostate HS (1)	0.33 hours	1.0 hours
Prostate MR (2)	0.66 hours	1.3 hours
Brain MR (3)	5.4 hours	14.8 hours